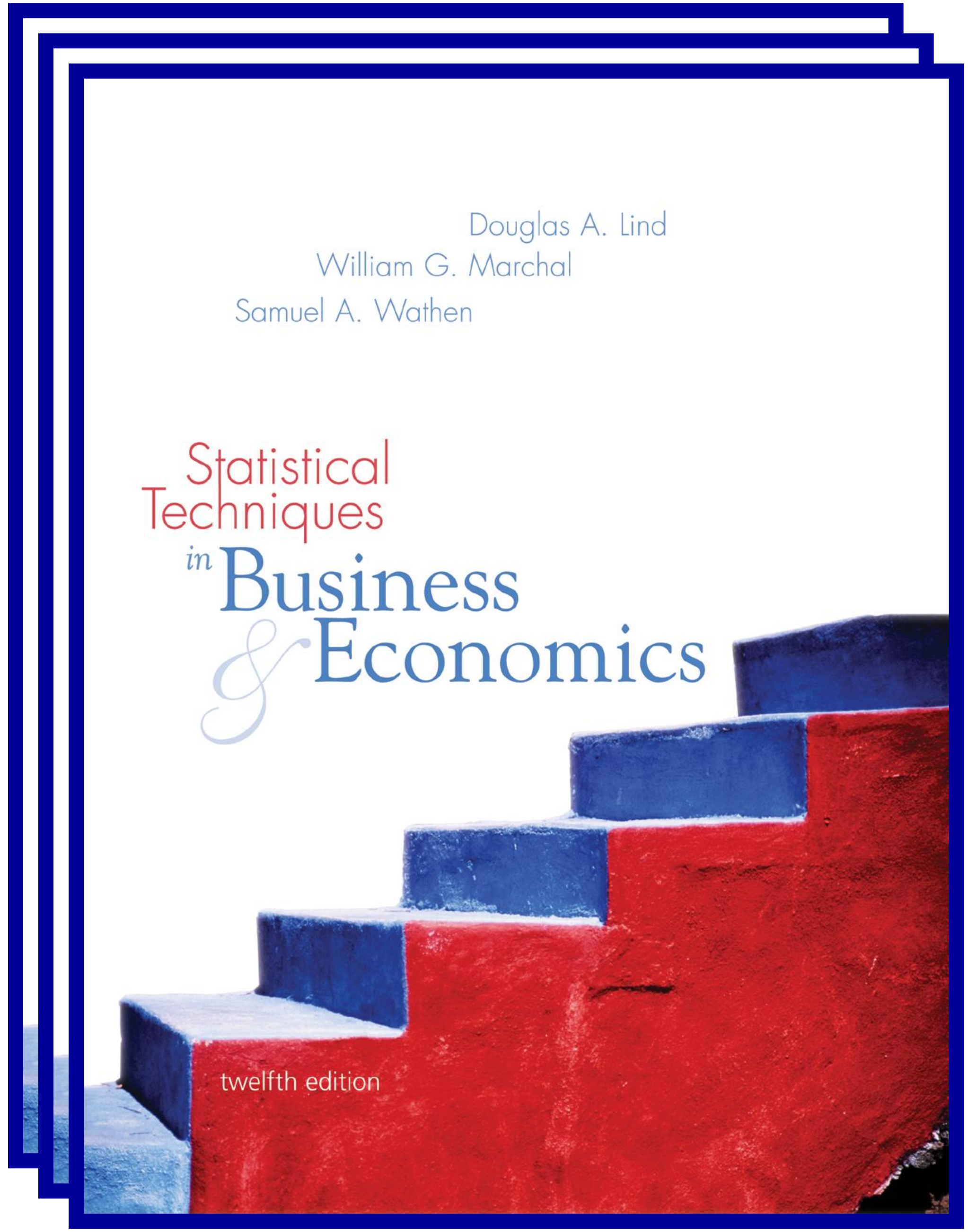
Chapter Three



Fb.com/groups/FCES.Mans

Chapter Three
Describing Data: Numerical Measures

GOALS

When you have completed this chapter, you will be able to:

ONE

Calculate the arithmetic mean, median, mode, weighted mean, and the geometric mean.

TWO

Explain the characteristics, uses, advantages, and disadvantages of each measure of location.

THREE

Identify the position of the arithmetic mean, median, and mode for both a symmetrical and a skewed distribution.

Goals

Chapter Three

Describing Data: Numerical Measures

FOUR

Compute and interpret the range, the mean deviation, the variance, and the standard deviation of ungrouped data.

FIVE

Explain the characteristics, uses, advantages, and disadvantages of each measure of dispersion.

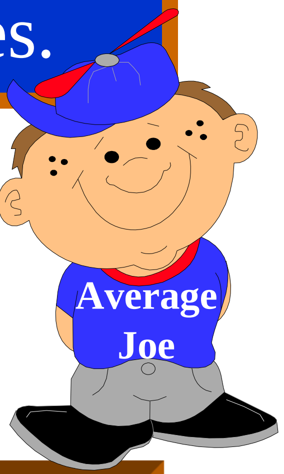
SIX

Understand Chebyshev's theorem and the Empirical Rule as they relate to a set of observations.

Goals

The **Arithmetic Mean** is the most widely used measure of location and shows the central value of the data.

It is calculated by summing the values and dividing by the number of values.



The major characteristics of the mean are:

- **It requires the interval scale.**
- **All values are used.**
- **It is unique.**
- **The sum of the deviations from the mean is 0.**

Characteristics of the Mean

For ungrouped data, the **Population Mean** is the sum of all the population values divided by the total number of population values:

$$\mu = \frac{\sum X}{N}$$

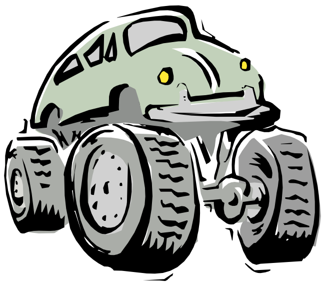
where

- μ is the population mean
- N is the total number of observations.
- X is a particular value.
- Σ indicates the operation of adding.

Population Mean

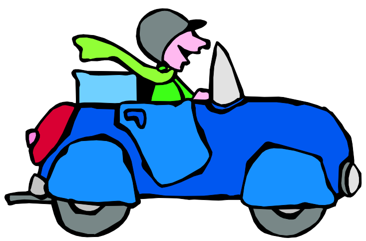
A **Parameter** is a measurable characteristic of a population.

The Kiers family owns four cars. The following is the current mileage on each of the four cars.



56,000

42,000



23,000

73,000

Find the mean mileage for the cars.

$$\mu = \frac{\sum X}{N} = \frac{56{,}000 + \ldots + 73{,}000}{4} = 48{,}500$$

Example 1

For ungrouped data, the sample mean is the sum of all the sample values divided by the number of sample values:

$$\bar{X} = \frac{\Sigma X}{n}$$

where *n* is the total number of values in the sample.

Sample Mean

A **statistic** is a measurable characteristic of a sample.

A sample of five executives received the following bonus last year ($000):

14.0,
15.0,
17.0,
16.0,
15.0

$$\overline{X} = \frac{\Sigma X}{n} = \frac{14.0 + \ldots + 15.0}{5} = \frac{77}{5} = 15.4$$

Example 2

Properties of the Arithmetic Mean

- Every set of interval-level and ratio-level data has a mean.
- All the values are included in computing the mean.
- A set of data has a unique mean.
- The mean is affected by unusually large or small data values.
- The arithmetic mean is the only measure of location where the sum of the deviations of each value from the mean is zero.

Consider the set of values: 3, 8, and 4. The **mean** is 5. Illustrating the fifth property

$$\Sigma(X-\bar{X}) = \left[(3-5)+(8-5)+(4-5)\right] = 0$$

Example 3

The **Weighted Mean** of a set of numbers X_1, X_2, ..., X_n, with corresponding weights w_1, w_2, ...,w_n, is computed from the following formula:

$$\bar{X}_w = \frac{(w_1 X_1 + w_2 X_2 + \ldots + w_n X_n)}{(w_1 + w_2 + \ldots w_n)}$$

During a one hour period on a hot Saturday afternoon cabana boy Chris served fifty drinks. He sold five drinks for $0.50, fifteen for $0.75, fifteen for $0.90, and fifteen for $1.10. Compute the weighted mean of the price of the drinks.

$$\bar{X}_w = \frac{5(\$0.50)+15(\$0.75)+15(\$0.90)+15(\$1.15)}{5+15+15+15}$$

$$= \frac{\$44.50}{50} = \$0.89$$

Example 4

The **Median** is the midpoint of the values after they have been ordered from the smallest to the largest.

There are as many values above the median as below it in the data array.

For an even set of values, the median will be the arithmetic average of the two middle numbers and is found at the (n+1)/2 ranked observation.

The ages for a sample of five college students are:
21, 25, 19, 20, 22.

Arranging the data in ascending order gives:

19, 20, 21, 22, 25.

Thus the median is 21.

The median (continued)

The heights of four basketball players, in inches, are: 76, 73, 80, 75.

Arranging the data in ascending order gives:

73, 75, 76, 80

Thus the median is 75.5.

The median is found at the $(n+1)/2 = (4+1)/2 = 2.5^{th}$ data point.

Example 5

Properties of the Median

- There is a unique median for each data set.
- It is not affected by extremely large or small values and is therefore a valuable measure of location when such values occur.
- It can be computed for ratio-level, interval-level, and ordinal-level data.
- It can be computed for an open-ended frequency distribution if the median does not lie in an open-ended class.

Properties of the Median

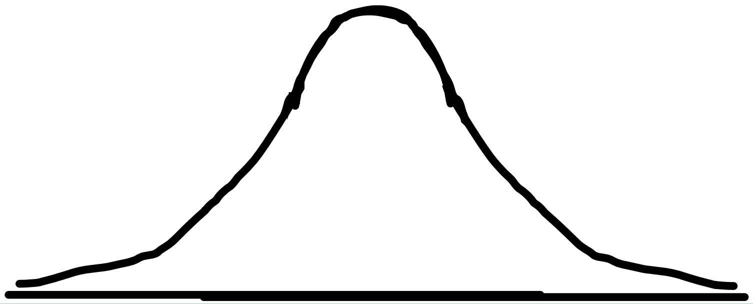
The **Mode** is another measure of location and represents the value of the observation that appears most frequently.

Example 6: The exam scores for ten students are: 81, 93, 84, 75, 68, 87, 81, 75, 81, 87. Because the score of 81 occurs the most often, it is the mode.

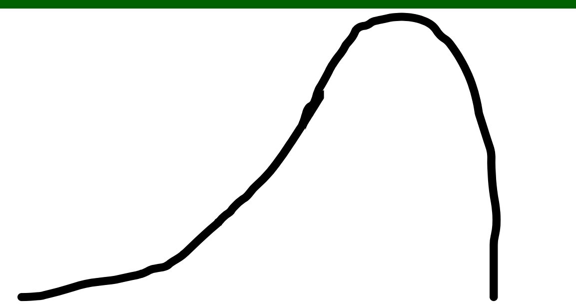
Data can have more than one mode. If it has two modes, it is referred to as bimodal, three modes, trimodal, and the like.

The Mode: Example 6

Symmetric distribution: **A distribution having the same shape on either side of the center**



Skewed distribution: One whose shapes on either side of the center differ; a nonsymmetrical distribution.

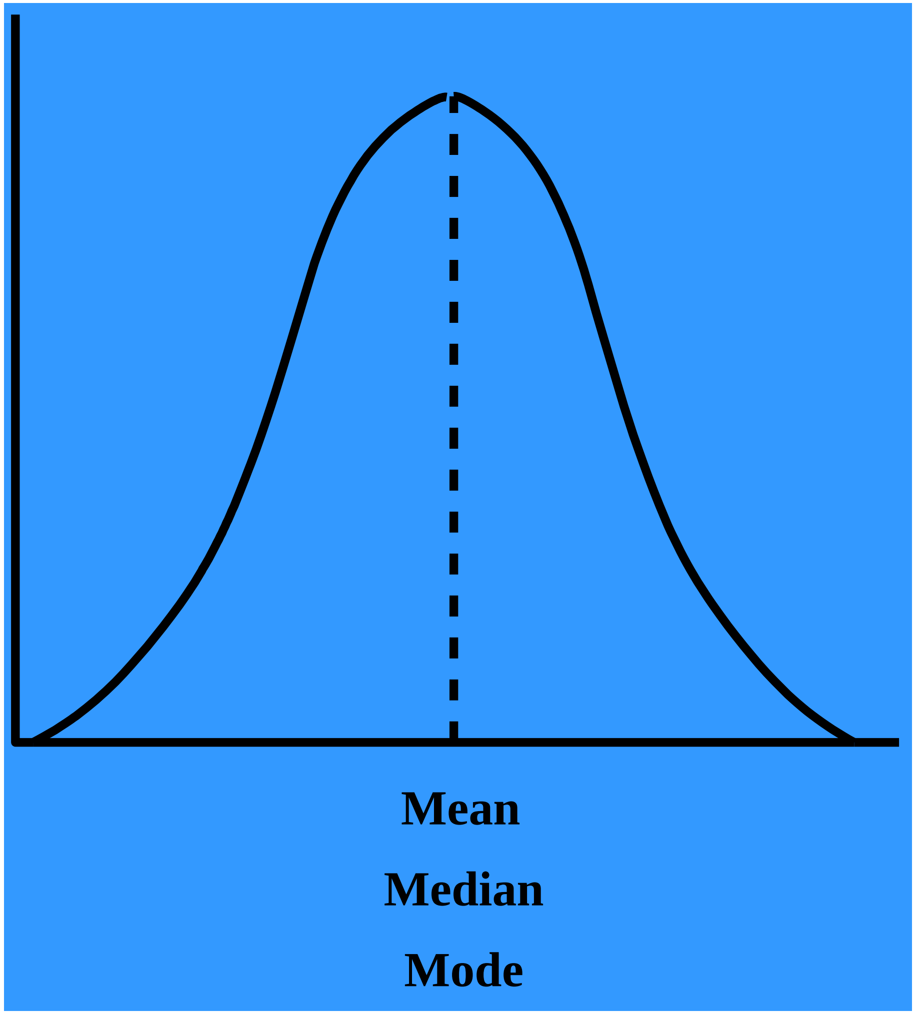


Can be positively or negatively skewed, or bimodal

The Relative Positions of the Mean, Median, and Mode

Zero skewness

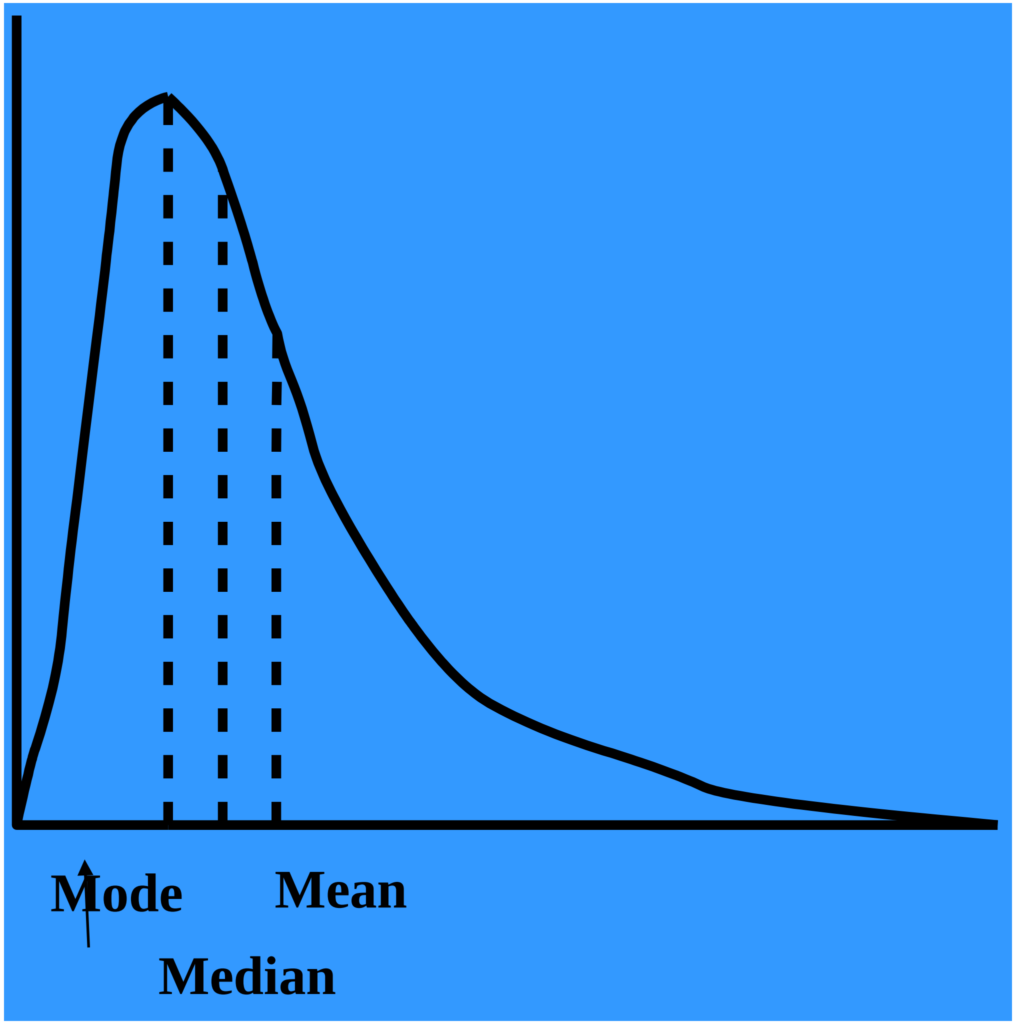
Mean
=Median
=Mode



The Relative Positions of the Mean, Median, and Mode:
Symmetric Distribution

- **Positively skewed**: Mean and median are to the right of the mode.

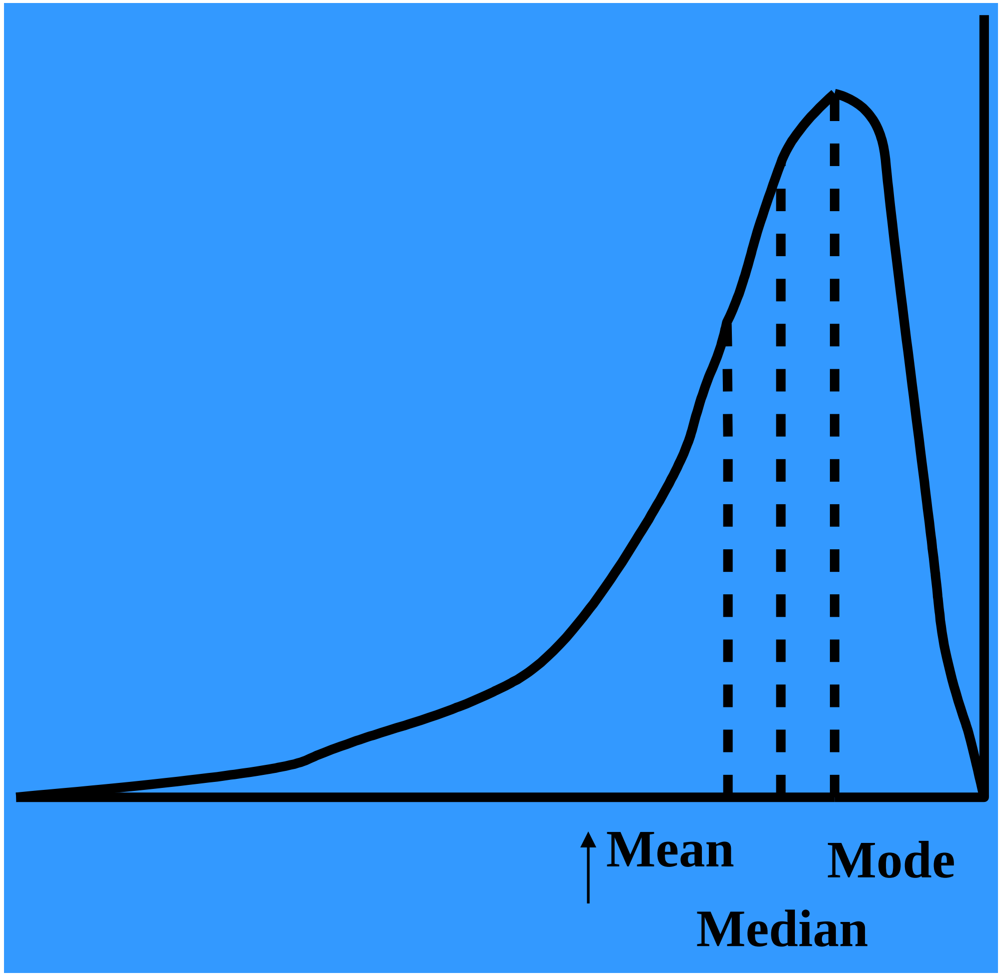
Mean>Median>Mode



The Relative Positions of the Mean, Median, and Mode: Right Skewed Distribution

Negatively Skewed: Mean and Median are to the left of the Mode.

Mean<Median<Mode



The Relative Positions of the Mean, Median, and Mode: Left Skewed Distribution

The Geometric Mean (*GM*) of a set of n numbers is defined as the *nth* root of the product of the *n* numbers. The formula is:

$$GM = \sqrt[n]{(X_1)(X_2)(X_3)\ldots(X_n)}$$

The geometric mean is used to average percents, indexes, and relatives.

The interest rate on three bonds were 5, 21, and 4 percent.

The arithmetic mean is (5+21+4)/3 =10.0.

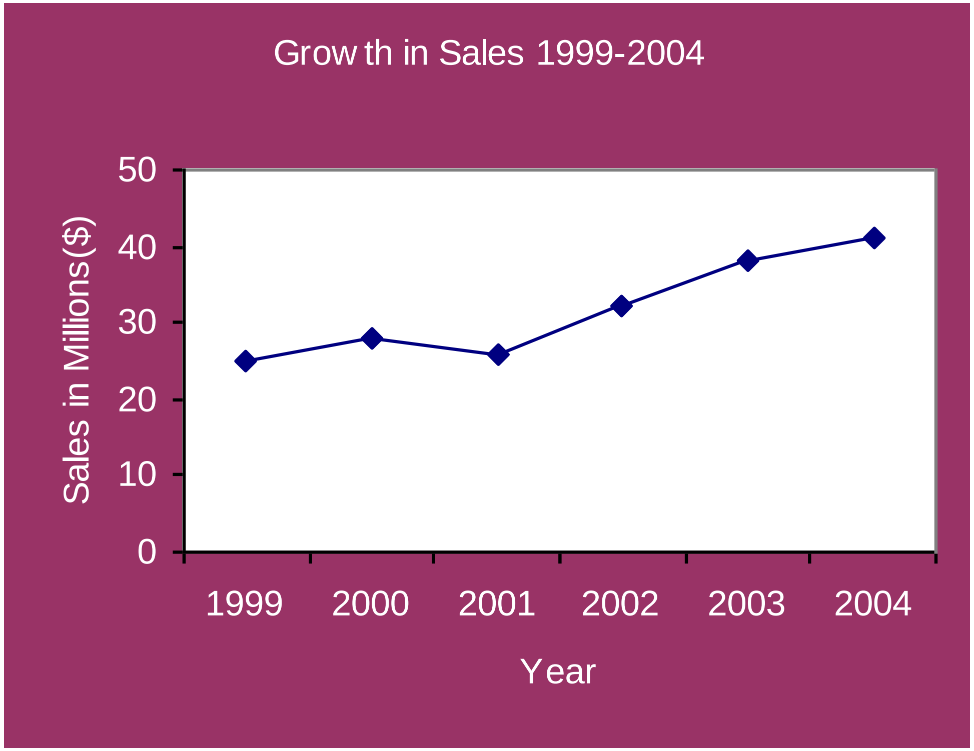
The geometric mean is

$$GM = \sqrt[3]{(5)(21)(4)} = 7.49$$

The *GM* gives a more conservative profit figure because it is not heavily weighted by the rate of 21percent.

Example 7

Another use of the geometric mean is to determine the percent increase in sales, production or other business or economic series from one time period to another.



$$GM = \sqrt[n]{\frac{(\text{Value at end of period})}{(\text{Value at beginning of period})}} - 1$$

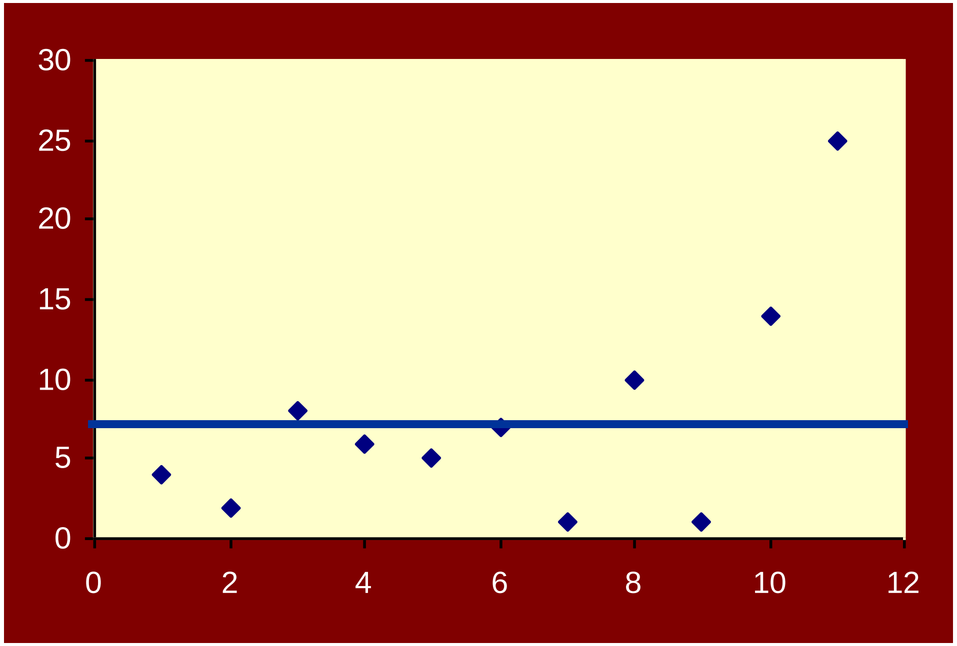
Geometric Mean continued

The total number of females enrolled in American colleges increased from 755,000 in 1992 to 835,000 in 2000. That is, the geometric mean rate of increase is 1.27%.

$$GM = \sqrt[8]{\frac{835{,}000}{755{,}000}} - 1 = .0127$$

Example 8

Dispersion refers to the spread or variability in the data.



Measures of dispersion include the following: range, mean deviation, variance, and standard deviation.

Range = Largest value – Smallest value

The following represents the current year's Return on Equity of the 25 companies in an investor's portfolio.

-8.1	3.2	5.9	8.1	12.3
-5.1	4.1	6.3	9.2	13.3
-3.1	4.6	7.9	9.5	14.0
-1.4	4.8	7.9	9.7	15.0
1.2	5.7	8.0	10.3	22.1

Highest value: 22.1

Lowest value: -8.1

Range = Highest value – lowest value
= 22.1-(-8.1)
= 30.2

Example 9

Mean Deviation

The arithmetic mean of the absolute values of the deviations from the arithmetic mean.

The main features of the mean deviation are:

- All values are used in the calculation.
- It is not unduly influenced by large or small values.
- The absolute values are difficult to manipulate.

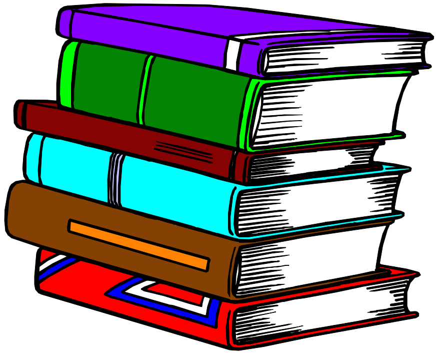
$$MD = \frac{\Sigma \left| X - \overline{X} \right|}{n}$$

The weights of a sample of crates containing books for the bookstore (in pounds) are:

103, 97, 101, 106, 103

Find the mean deviation.

$\overline{X} = 102$



The mean deviation is:

$$MD = \frac{\Sigma\left|X - \overline{X}\right|}{n} = \frac{\left|103 - 102\right| + \ldots + \left|103 - 102\right|}{5}$$

$$= \frac{1 + 5 + 1 + 4 + 5}{5} = 2.4$$

Example 10

Variance: the arithmetic mean of the squared deviations from the mean.

Standard deviation: The square root of the variance.

Variance and standard Deviation

The major characteristics of the **Population Variance** are:

Not influenced by extreme values.

- The units are awkward, the square of the original units.
- All values are used in the calculation.

Population Variance formula:

$$\sigma^2 = \frac{\Sigma (X - \mu)^2}{N}$$

X is the value of an observation in the population

m is the arithmetic mean of the population

N is the number of observations in the population

Population Standard Deviation formula:

$$\sigma = \sqrt{\sigma^2}$$

Variance and standard deviation

In Example 9, the variance and standard deviation are:

$$\sigma^2 = \frac{\Sigma (X - \mu)^2}{N}$$

$$\sigma^2 = \frac{(-8.1-6.62)^2 + (-5.1-6.62)^2 + \ldots + (22.1-6.62)^2}{25}$$

$$\sigma^2 = 42.227$$

$$\sigma = 6.498$$

Example 9 continued

Sample variance (s^2)

$$s^2 = \frac{\Sigma(X - \bar{X})^2}{n-1}$$

Sample standard deviation (s)

$$s = \sqrt{s^2}$$

Sample variance and standard deviation

The hourly wages earned by a sample of five students are:

$7, $5, $11, $8, $6.

Find the sample variance and standard deviation.

$$\bar{X} = \frac{\Sigma X}{n} = \frac{37}{5} = 7.40$$

$$s^2 = \frac{\Sigma\left(X - \bar{X}\right)^2}{n-1} = \frac{(7-7.4)^2 + \ldots + (6-7.4)^2}{5-1}$$
$$= \frac{21.2}{5-1} = 5.30$$

$$s = \sqrt{s^2} = \sqrt{5.30} = 2.30$$

Example 11

Chebyshev's theorem: For any set of observations, the minimum proportion of the values that lie within *k* standard deviations of the mean is at least:

$$1-\frac{1}{k^2}$$

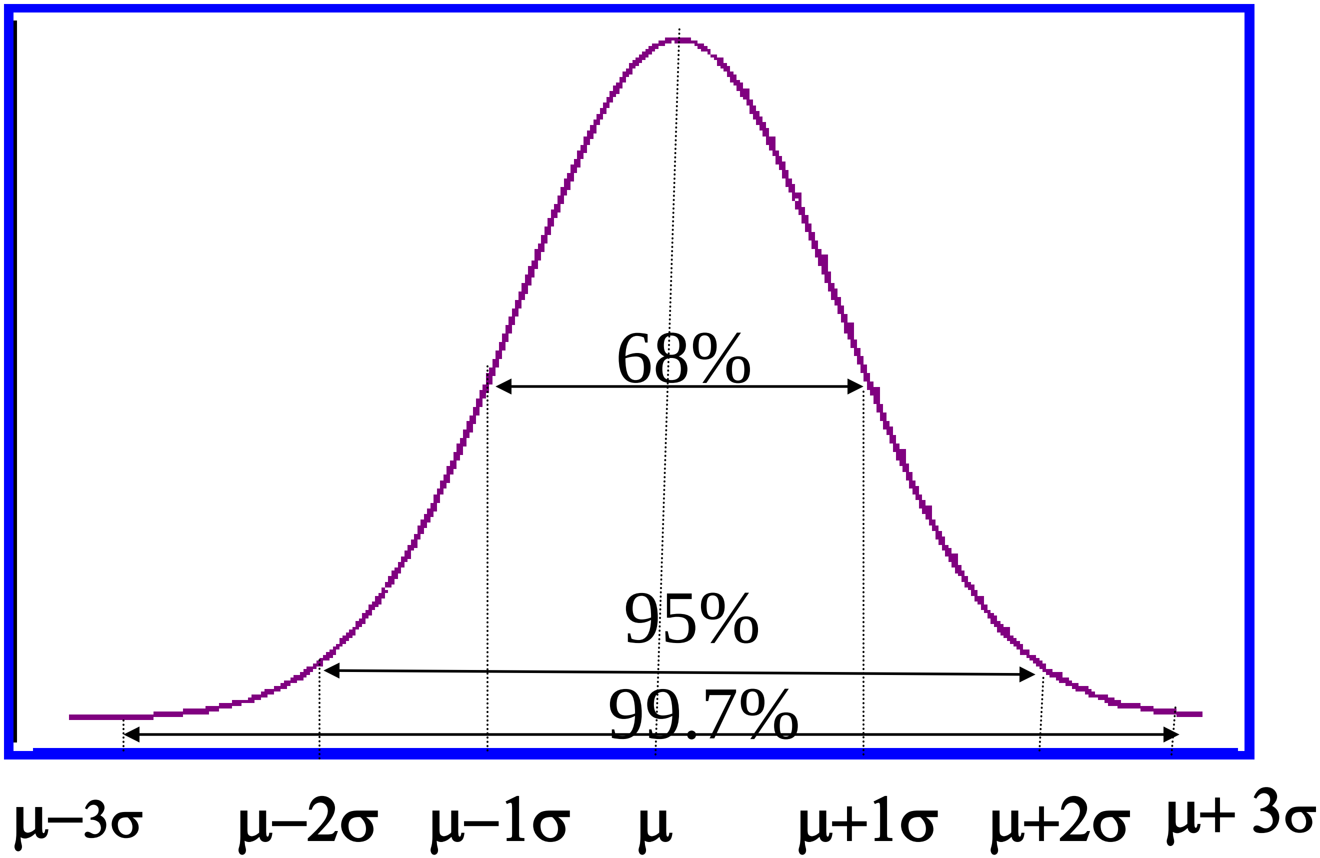
- where *k* is any constant greater than 1.

Chebyshev's theorem

Empirical Rule: For any symmetrical, bell-shaped distribution:

- About 68% of the observations will lie within 1*s* the mean
- About 95% of the observations will lie within 2*s* of the mean
- Virtually all the observations will be within 3*s* of the mean

Bell-Shaped Curve showing the relationship between σ and μ.



- The **Mean** of a sample of data organized in a frequency distribution is computed by the following formula:

$$\overline{X} = \frac{\Sigma Xf}{n}$$

The Mean of Grouped Data

A sample of ten movie theaters in a large metropolitan area tallied the total number of movies showing last week. Compute the mean number of movies showing.

Movies showing	frequency f	class midpoint X	$(f)(X)$
1 up to 3	1	2	2
3 up to 5	2	4	8
5 up to 7	3	6	18
7 up to 9	1	8	8
9 up to 11	3	10	30
Total	10		66

$$\overline{X} = \frac{\Sigma X}{n} = \frac{66}{10} = 6.6$$

Example 12

The **Median** of a sample of data organized in a frequency distribution is computed by:

$$Median = L + \frac{\frac{n}{2} - CF}{f}(i)$$

where L is the lower limit of the median class, CF is the cumulative frequency preceding the median class, f is the frequency of the median class, and i is the median class interval.

The Median of Grouped Data

To determine the median class for grouped data

Construct a cumulative frequency distribution.

Divide the total number of data values by 2.

Determine which class will contain this value. For example, if n=50, 50/2 = 25, then determine which class will contain the 25th value.

Finding the Median Class

Movies showing	Frequency	Cumulative Frequency
1 up to 3	1	1
3 up to 5	2	3
5 up to 7	3	6
7 up to 9	1	7
9 up to 11	3	10

Example 12 continued

From the table, L=5, n=10, f=3, i=2, CF=3

$$Median = L + \frac{\frac{n}{2} - CF}{f}(i) = 5 + \frac{\frac{10}{2} - 3}{3}(2) = 6.33$$

Example 12 continued

The **Mode** for grouped data is approximated by the midpoint of the class with the largest class frequency.

The modes in example 12 are 6 and 10 and so is bimodal.

The Mode of Grouped Data